ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

# Non Stationary Signals (Voice) Verification System Using Wavelet Transform

**PPS Subhashini**
**Associate Professor,**
**Department of ECE,**
**RVR & JC College of Engineering, Guntur.**

**Dr.M.Satya Sairam**
**Professor & HOD,**
**Department of ECE,**
**Chalapathi Institute of Engineering and Technology, Guntur.**

**Dr.D.Srinivasa Rao**
**Professor & HOD,**
**Department of ECE,**
**J N T U H, Hyderabad.**

**Abstract:**
Speech recognition (SR) is the inter-disciplinary sub-field of computational linguistics which incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields to develop methodologies and technologies that enables the recognition and translation of spoken language into text by computers and computerized devices such as those categorized as smart technologies and robotics. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT), or Voice Verification. This document looks at a new technique for analyzing and compressing speech signals using wavelets. Very simply wavelets are mathematical functions of finite duration with an average value of zero that are useful in representing data or other functions.

The implemented voice recognition system is word dependent voice verification system combining the RASTA and LPC. The voice signal is filtered using the special purpose voice signal filter using the Relative Spectral Algorithm (RASTA). The efficiency of the system can be improved using Hidden Markov Model (HMM) technique. In speech recognition, the HMM model optimizes the probability of the training set to detect a particular speech. The probability function is performed by the Viterbi algorithm. This algorithm is a procedure used to determine an optimal state sequence from a given observation sequence.

**Keywords:**
Speech Recognition, Voice verification, Wavelets, LPC, RASTA, HMM.

**Introduction:**
Speech is a very basic way for humans to convey information to one another. With a bandwidth of only 4 kHz, speech can convey information with the emotion of a human voice. People want to be able to hear someone's voice from anywhere in the world. As if the person was in the same room. As a result a greater emphasis is being placed on the design of new and efficient speech codes for voice communication and transmission. Today applications of speech coding and compression have become very numerous. Many applications involve the real time coding of speech signals, for use in mobile satellite communications, cellular telephony, and audio for videophones or video teleconferencing systems. Other applications include the storage of speech for speech synthesis and playback, or for the transmission of voice at a later time.

Some examples include voice mail systems, voice memo wristwatches, voice logging recorders and interactive PC software. Traditionally speech coders can be classified into two categories: waveform coders and analysis/synthesis vocoders (from .voice coders.). Waveform coders attempt to copy the actual shape of the signal produced by the microphone and its associated analogue circuits. A popular waveform coding technique is pulse code modulation (PCM). This is used in telephony today. Vocoders use an entirely different approach to speech coding, known as parameter coding, or analysis/synthesis coding where no attempt is made at reproducing the exact speech waveform at the receiver, only a signal perceptually equivalent to it.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

These systems provide much lower data rates by using a functional model of the human speaking mechanism at the receiver. One of the most popular techniques for analysis synthesis coding of speech is called Linear Predictive Coding (LPC). Some higher quality vocoders include RELP (Residual Excited Linear Prediction) and CELP (Code Excited Linear Prediction).Any signal can be represented by a set of scaled and translated versions of a basic function called the mother wavelet. This set of wavelet functions forms the wavelet coefficients at different scales and positions and results from taking the wavelet transform of the original signal. The coefficients represent the signal in the wavelet domain and all data operations can be performed using just the corresponding wavelet coefficients.

Speech is a non-stationary random process due to the time varying nature of the human speech production system. Non-stationary signals are characterized by numerous transitory drifts, trends and abrupt changes. The localization feature of wavelets, along with its time-frequency resolution properties makes them well suited for coding speech signals. In designing a wavelet based speech coder, the major issues explored in this thesis are:

i. Choosing optimal wavelets for speech,

ii. Decomposition level in wavelet transforms,

iii. Threshold criteria for the truncation of coefficients,

iv. Efficiently representing zero valued coefficients and

v. Quantizing and digitally encoding the coefficients.

The performance of the wavelet compression scheme in coding speech signals and the Quality of the reconstructed signals is also evaluated. Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker. Fig.1 shows the basic structures of speaker identification and verification systems.

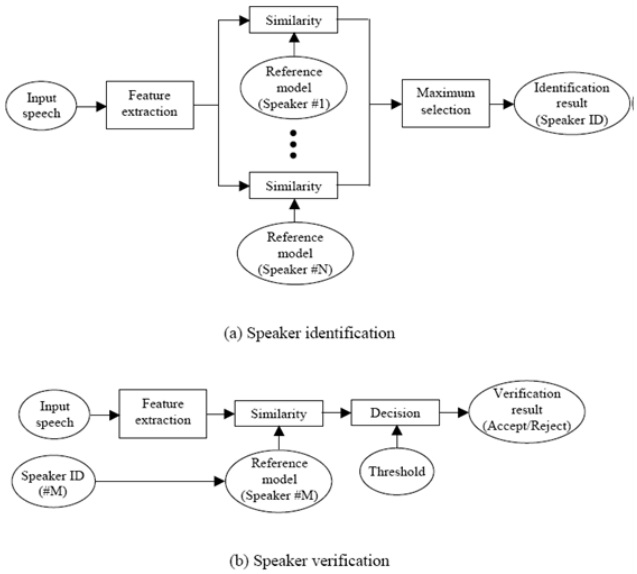


**Fig. 1 Basic structures of speaker recognition system**

## Wave Lets:

The fundamental idea behind wavelets is to analyze according to scale. The wavelet analysis procedure is to adopt a wavelet prototype function called an analyzing wavelet or mother wavelet. Any signal can then be represented by translated and scaled versions of the mother wavelet. Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques such as Fourier analysis miss aspects like trends, breakdown points, discontinuities in higher derivatives, and self-similarity. Furthermore, because it affords a different view of data than those presented by traditional techniques, it can compress or de-noise a signal without appreciable degradation. The fig2 below illustrates four different types of wavelet basis functions.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

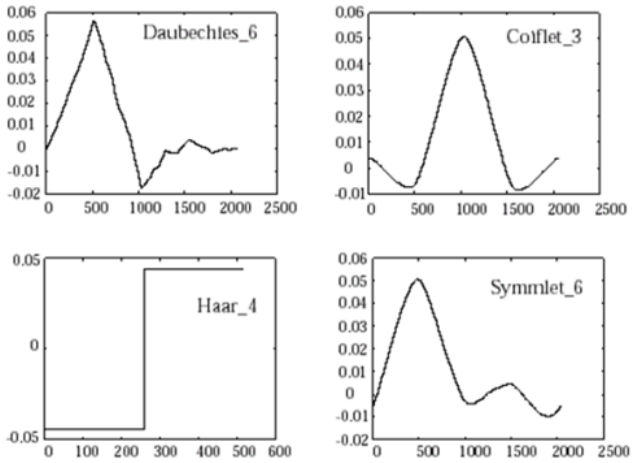


**Fig. 2 Different Wavelet Families**

The different families make trade-offs between how compactly the basic functions are localized in space and how smooth they are. Within each family of wavelets (such as the Daubechies family) are wavelet subclasses distinguished by the number of filter coefficients and the level of iteration. Wavelets are most often classified within a family by the number of vanishing moments. This is an extra set of mathematical relationships for the coefficients that must be satisfied. The extent of compactness of signals depends on the number of vanishing moments of the wavelet function used. A more detailed discussion is provided in the next section.

## VOICE SIGNAL ANALYSIS:

### RASTA (Relative Spectral Algorithm):

RASTA or Relative Spectral Algorithm as it is known is a technique that is developed as the initial stage for voice recognition. This method works by applying a band-pass filter to the energy in each frequency sub-band in order to smooth over short-term noise variations and to remove any constant offset. Invoice signals, stationary noises are often detected. Stationary noises are noises that are present for the full period of a certain signal and does not have diminishing feature . Their property does not change over time. The assumption that needs to be made is that the noise varies slowly with respect to speech. This makes the RASTA a perfect tool to be included in the initial stages of voice signal filtering to remove stationary noises . The stationary noises that are identified are noises in the frequency range of1Hz - 100Hz.

### LPC(Linear Predictive Coding):

The LPC will analyze the signal by estimating or predicting the formants. Then, the formants effects are removed from the speech signal. The intensity and frequency of the remaining buzz is estimated. So by removing the formants from the voices signal will enable us to eliminate the resonance effect. This process is called inverse filtering. The remaining signal after the formant has been removed is called the residue. In order to estimate the formants, coefficients of the LPC are needed. The coefficients are estimated by taking the mean square error between the predicted signal and the original signal. By minimizing the error, the coefficients are detected with a higher accuracy and the formants of the voice signal are obtained. Formant is one of the major components of speech. The frequencies at which the resonant peaks occur are called the formant frequencies or simply formants .The formant of the signal can be obtained by analyzing the vocal tract frequency response.

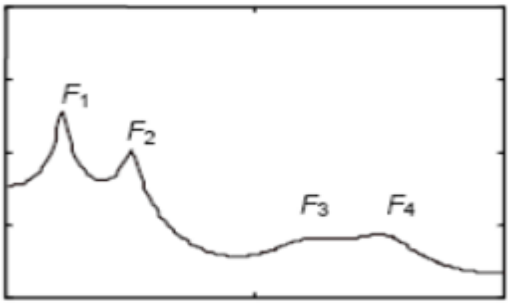


**Fig.3 Vocal tract frequency response**

Fig.3shows the vocal tract frequency response. The x-axis represents the frequency scale and the y-axis represents the magnitude of the signal. As it can be seen, the formants of the signals are classified as F1,F2, F3 and F4. Typically a voice signal will contain three to five formants. But in most voice signals, up to four formants can be detected. In order to obtain the formant of the voice signals, the LPC (Linear

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Predictive Coding) method is used. The LPC (Linear Predictive Coding) method is derived from the word linear prediction. Linear prediction as the term implies is a type of mathematical operation. This mathematical function which is used in discrete time signal estimates the future values based upon a linear function of previous samples.

$$\overset{\wedge}{x}(n) = -\sum_{i=1}^{p} a(i)x(n-i)$$

Where $\hat{x}(n)$ is the predicted or estimated value and $x(n-i)$ is the pervious value. By expanding this equation

$$\overset{\wedge}{x}(n) = -a(1)x(n-1) - a(2)x(n-2) - a(3)x(n-3)$$

## System Implementation:

In order to implement the system, a certain methodology is implemented by decomposing the voice signal to its approximation and detail. From the approximation and detail coefficients that are extracted, the methodology is implemented in order to carry out the recognition process. The proposed methodology for the recognition phase is the statistical calculation. Four different types of statistical calculations are carried out on the coefficients. The statistical calculations that are carried out are mean, standard deviation, variance and mean of absolute deviation. The wavelet that is used for the system is thesymlet 7 wavelet as that this wavelet has a very close correlation with the voice signal. This is determined through numerous trial and errors. The coefficients that are extracted from the wavelet decomposition process is the second level coefficients as the level two coefficients contain most of the correlated data of the voice signal. The data at higher levels contains very little amount of data deeming it unusable for there cognition phase. Hence for initial system implementation, the level two coefficients are used. The coefficients are further threshold to remove the low correlation values, and using this coefficients statistical computation is carried out. The statistical computation of the coefficients is used in comparison of voice signal together with the formant estimation and the wavelet energy. All the extracted information acts like a 'fingerprint' for the voice signals. The percentage of verification is calculated by comparing the current values signal values against the registered voice signal values. The percentage of verification is given by: Verification % = (Test value / Registered value) x100Between the tested and registered value, which ever value is higher is taken as the denominator and the lower value is taken as the numerator. Fig. below shows the complete flowchart which includes all the important system components that are used in the voice verification program.

## Flow Chart:

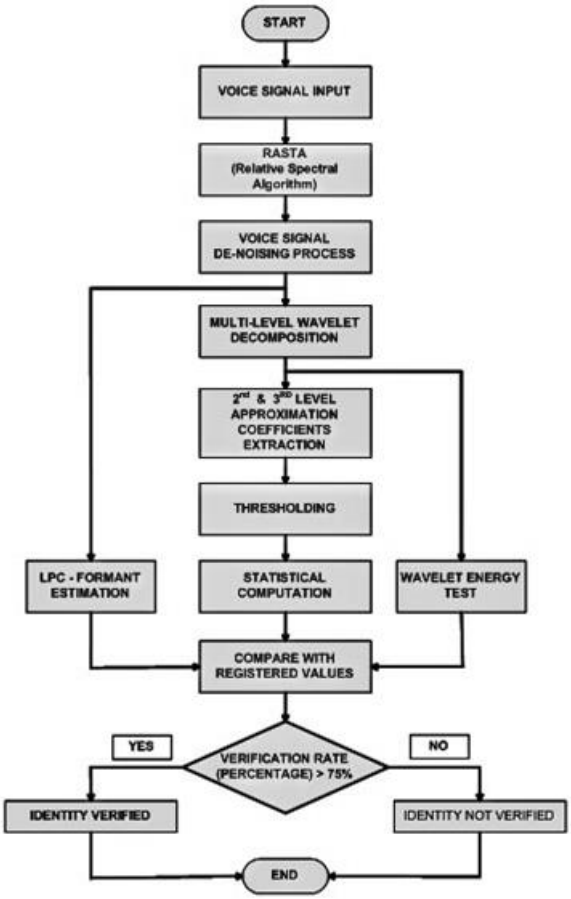


Fig.4 Complete System Flowchart

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

## GRAPHICAL USER INTERFACE:

Fig.5below shows the GUI implementation for the Voice Registration Section. This GUI enables the user to register an individual's voice signal using the pre-loaded voice signals that are saved in the program. The STEP 1 panel shows the pre-loaded voice signals contained in the program. The voice signals are obtained from a clean and noise-free environment. The user can select the available voice signal from the popup menu. The Show Voice Plot button enables the user to view the voice plot in the graph shown in the panel. The STEP 2 panel contains the function that enables the user to de-noise the signal. By pressing the De-Noise button, the user will be able to de-noise the signal and view the de-noised signal in the graph shown in the plot. The Extract Coefficients button enables the user to view the DWT Coefficients Detail and Approximation plot. The STEP 3 panel shows the extracted coefficients plot which is the 2nd level approximation& detail and the 3rdlevel approximation & detail. The STEP 4 panel shows the recognition methodology of the program. The Compute Values button performs the statistical computation on the 2nd level approximation and the 3rd level approximation and displays these values. At the same time the formant values and the wavelet energy of the voice signal is calculated and shown.

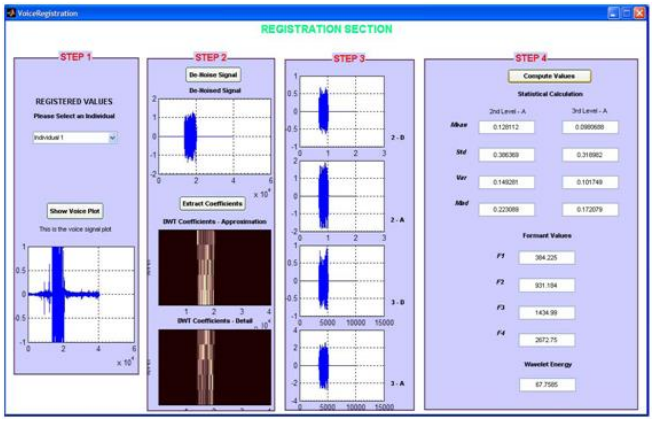


**Fig.5Voice Registration GUI**

Fig.6 below shows the GUI (Graphical User Interface)implementation for the Voice Verification Section.This GUI enables the user to verify an individual'svoice signal using the pre-loaded voice signals that aresaved in the program.

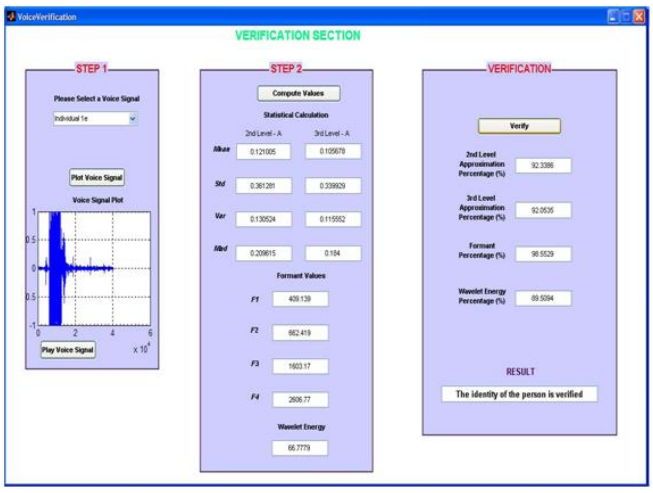


**Fig. 6 Voice Verification GUI**

The STEP 1 panel shows the pre-loaded voice signals contained in the program. These voice signals are recorded from different individuals saying their own name. The user can select the available voice signal from the pop-up menu. The Plot Voice Signal button enables the user to view the voice plot in the graph shown in the panel. The STEP 2 panel shows the recognition methodology of the program. The Compute Values button performs the statistical computation of the 2nd level approximation and the 3rdlevel approximation and displays these values. At the same time the formant values and the wavelet energy of the voice signal is calculated and shown. The VERIFICATION panel shows the verification process of the system. The overall percentage values of the statistical computation, formant values and the wavelet energy are displayed.

## RESULTS:

## TABLE 1 COMPARISON TEST 1

| Individual | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | Verified | Not Verified | Not Verified | Not Verified | Not Verified |
| 2 | Not Verified | Verified | Not Verified | Not Verified | Not Verified |
| 3 | Not Verified | Not Verified | Verified | Not Verified | Not Verified |
| 4 | Not Verified | Not Verified | Not Verified | | Not Verified |
| 5 | Not Verified | Not Verified | Not Verified | Not Verified | Verified |

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

**TABLE 2 COMPARISON TEST 2**

| Individual | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | Verified | Not Verified | Not Verified | Not Verified | Not Verified |
| 2 | Not Verified | | Not Verified | Not Verified | Not Verified |
| 3 | Not Verified | Not Verified | Verified | Not Verified | Not Verified |
| 4 | Not Verified | Not Verified | Not Verified | Verified | Not Verified |
| 5 | Not Verified | Not Verified | Not Verified | Not Verified | Verified |

**TABLE 3 COMPARISON TEST 3**

| Individual | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | Verified | Not Verified | Not Verified | Not Verified | Not Verified |
| 2 | Not Verified | Verified | Not Verified | Not Verified | Not Verified |
| 3 | Not Verified | Not Verified | Verified | Not Verified | Not Verified |
| 4 | Not Verified | Not Verified | Not Verified | Verified | Not Verified |
| 5 | Not Verified | Not Verified | Not Verified | Not Verified | Verified |

**TABLE 4 COMPARISON TEST 4**

| Individual | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | Verified | Not Verified | Not Verified | Not Verified | Not Verified |
| 2 | Not Verified | Verified | Not Verified | Not Verified | Not Verified |
| 3 | Not Verified | Not Verified | | Not Verified | Not Verified |
| 4 | Not Verified | Not Verified | Not Verified | Verified | Not Verified |
| 5 | Not Verified | Not Verified | Not Verified | Not Verified | Verified |

**TABLE 5 COMPARISON TEST 5**

| Individual | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | Verified | Not Verified | Not Verified | Not Verified | Not Verified |
| 2 | Not Verified | Verified | Not Verified | Not Verified | Not Verified |
| 3 | Not Verified | Not Verified | Verified | Not Verified | Not Verified |
| 4 | Not Verified | Not Verified | Not Verified | | Not Verified |
| 5 | Not Verified | Not Verified | Not Verified | Not Verified | Verified |

From the Tables above of the verification result shows from the five random tests carried out, at anyone given time, the program can successfully verify 4out 5 persons accurately. The complete systems which constitutes all the system components for the recognition methodology is one of the main reasons for the high accuracy of the system. Currently, the percentage of verification is set at an average value of78.75%. The verification rate can be further increased or decreased by adjusting the percentage of verification to a higher or lower value. By substituting a lower value, the system will be less secure while a higher value could jeopardize the accessibility rate of the system because of the certain level of tolerance is required for the voice signal as it tends to change with internal and external factors.

## Conclusion:

The Voice Recognition Using Wavelet Feature Extraction employ wavelets in voice recognition for studying the dynamic properties and characteristics of the voice signal. This is carried out by estimating the formant and detecting the pitch of the voice signal by using LPC (Linear Predictive Coding). The voice recognition system that is developed is word dependant voice verification system used to verify the identity of an individual based on their own voice signal using the statistical computation, formant estimation and wavelet energy. A GUI is built to enable the user to have an easier approach in observing the step-by-step process that takes place in Wavelet Transform. By using the fifty preloaded voice signals from five individuals, the verification tests have been carried and an accuracy rate of approximately 80 % has been achieved. The system can be enhanced further by using advanced pattern recognition techniques such as Neural Network or Hidden Markov Model (HMM).

## References:

[1]Rajparthiban Kumar, C. V. Aravind , Kanendra Naidu and Anis Fariza, “Development of Novel Voice Verification system Using Wavelets”, IEEE International conference on Computer and Communication Engineering , Vol. III, 2008.

[2]Rasta-PLP Speech analysis, Hermansky, H., Morgan, N., Bayya, A, and Kohn, P. ICSI Technical Report TR-91-069, Berkeley, California.

[3]http://www.learnartificialneuralnetworks.com/back propagation. html

[4]Oppenheim, A.V. and Shafer, R, “Discrete-Time Signal Processing” Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.

[5]Bao Liu, Sherman Riemenschneider, and ZuoweiShen, “An Adaptive Time-Frequency Representation and its Fast Implementation”, ASME

Journal of Vibration and Acoustics, Volume 129 (2007), Issue 2, pp. 169-178.

[6]Sailaja Mungamuri & P.P.S. Subhashini, Text Dependent Speaker Recognition Using RASTA LPC and Discrete Wavelet Transform, IJMETMR, Volume No: 2 (2015), Issue No: 7 (July), http://www.ijmetmr.com/oljuly2015/SailajaMungamuri-PPSSubhashini-42.pdf

[7]Amara Graps, "An Introduction to Wavelets," IEEE Computational Science and Engineering, vol. 2, no. 2, pp. 50-61,summer, 1995.

[8]M. Vetterli and C. Herley, "Wavelets and filter banks theory and design", IEEE Trans. on Signal Proc, Sep. 1992.

[9]S.Akarsh & K.Avinash, Speech Processing, Speech Synthesis & Speech Recognition, May 2010, http://www.yuvaengineers.com/speech-processing-speech-synthesis-speech-recognition-s-akarsh-k-avinash/

[10]Giuliano Antoniol, Vincenzo Fabio Rollo, Gabriele Venturi, IEEE Transactions on Software Engineering, LPC &Cepstrum coefficients for Mining Time Variant Information from Software Repositories, University Of Sannio, Italy